Genomics 94 (2009) 363–368

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Minireview

Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing

Daniel Summerer *

febit biomed gmbh, Im Neuenheimer Feld 519, 69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 13 July 2009
Accepted 22 August 2009
Available online 29 August 2009

Keywords:
Next-generation-sequencing
Sequence capture
Microarrays
Hybridzation probes

ABSTRACT

Next-generation sequencing has still not reached its full potential due to the technical inability of effectively targeting desired genomic regions of interest. Once available, methods adressing this bottleneck will dramatically reduce cost and enable the efficient analysis of complex samples.
Recently, a number of possible approaches for genomic-scale sequence enrichment have been reported using different strategies. All methods basically rely on sequence-specific nucleic acid hybridization, however, they differ in several aspects such as the use of solid phase versus solution phase hybridization, probe design and overall workflows with implications for automation.
Overall, several key challenges of genome-wide sequence enrichment have become clear after these studies that remain to be overcome. We summarize the different technologies and highlight individual characteristics related to general potential and different suitabilities for specific applications.

Introduction

The vast capacity of next-generation sequencers (NGS) has tremendously increased the scope and comprehensiveness of genomics projects [1-6]. Pro- and eukaryotic genomes are now accessible within days or weeks, fundamentally changing typical project scales in genetics research. Besides the effect of increased throughput, large-scale sequencing studies are now open to many more research laboratories from different disciplines due to associated reduction of cost. This democratizing effect of NGS technologies will accelerate new discoveries and promises to diversify the way in which genetic studies are designed.

However, the enhanced throughput on the sequencing side has not been flanked by the development of suited sample preparation techniques allowing for the focussed analysis of genomic subsets [7]. In fact, until very recently, no efficient methods have been available to enrich DNA sequences out of complex genomic mixtures at a capacity exceeding the low kilobase range of classic PCR. Though PCR as a well-established method of enrichment is basically feasible, its sequence scale and level of multiplexing rather match the throughput of traditional sanger sequencing. The megabase capacity of NGS instruments however may require thousands of PCR reactions for a typical study, including potential optimization of individual reactions, synthesis of primer pairs and normalization. This lack of large scale enrichment methods represents a serious bottleneck for the exploitation of NGS instruments full potential.

The most immediate need for sequence enrichment originates from the current capacities of NGS platforms that do not allow sequencing of whole genomes of complex eukaryotic organisms with reasonable effort [8,9]. This essentially prevents the advantages of NGS for studies involving human and many eukaryotic model organisms. Additionally, large-scale enrichment methods might well play their part even after the next leap in sequencing throughput. Though complete sequencing of a human genome in one instrument run at a cost of about $1000 is certainly a next milestone of DNA sequencing technology [10], a further level is the multiplexing of several genomes within one run [11,12]. Beyond that, applying DNA sequencing to even more complex samples, for example in human population studies, analysis of microbial communities, host–pathogen mixtures or somatic variants might again substantially benefit from sequence enrichment methods of suited efficiency and scale [13]. Since even large regions of interest like a whole exome typically represent only a few percent of a genome for human and many model organisms, efficient targeted sequencing can dramatically reduce cost and effort. This reduction becomes even larger with multi-genome complexities of the analyzed sample. From a practical point of view, sequence enrichment methods also enable the efficient use of the in-built compartments of current NGS platforms for multiplexing of several, separated samples without indexing strategies.

Recently, a number of approaches have been reported that might help to overcome these current bottlenecks [14-22]. All of these rely on complementary hybridization of nucleic acid capture probes to the targeted DNA sequences. However, there are also substantial differences. Some methods use solution phase and others solid phase hybridization and the methods differ in overall workflows and ease of automation. The design of the individual sequence capture

* Fax: +49 6221 6510 390.
E-mail address: daniel.summerer@febit.de.

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.08.012

steps additionally influences the accessibility for different types of target regions and overall efficiency. Here, we summarize these recent developments in this highly dynamic and open field and point out similarities and substantial differences between the approaches. We thereby emphasize technological and conceptual differences rather than absolute performance parameters since experiments for direct comparison are not available so far.

Critical parameters of genomic-scale sequence enrichment

It has become clear from recent studies that efficient capture of target sequences on a genomic scale imposes several special requirements on respective enrichment technologies.

For example, many relevant regions in targeted NGS, such as exons, vary in size and sequence properties and are discontiguously distributed in the genome within a context of low complexity sequence. An ideal sequence enrichment method should therefore allow random access to multiple different loci relatively independent of their size, sequence composition and spatial distribution. This must be achieved at a multiplexing grade that matches NGS capacities. Individual loci should not only be generally accessible for capture but should be enriched with equal efficiencies to allow for complete and uniform coverage of the targeted region. This is essential for economic enrichment, since a high uniformity avoids redundant reads from overcaptured regions. In terms of enrichment performance, an ideal enrichment method would thus allow for complete and entirely even coverage of the target region with the minimum depth required for reliable nucleotide calling. This has to be achieved without introduction of bias into allele representations. For economic reasons, enrichment efficiency should finally be such that data output of the NGS intrument is only related to target region with minimal background sequence data.

On a molecular level, several of these performance parameters seem to largely depend on two basic process properties. Firstly, the hybridization step itself determines specificity and uniformity of binding capacity for individual target regions and allelic variants and thus influences the enrichment efficiency and consequently the fraction of on-target reads in the NGS instruments output. Good performance and high sequence capacity of the hybridzation step can be adressed by using specific hybridization probe libraries with sizes matching NGS scale. Established design rules for genomic DNA microarray hybridization, e.g., aiming at maximal binding specificity, similar melting temperatures and minimal content of low complexity sequence have thereby been applied.

Secondly, the individual methods have different local coverage distributions at specific target loci as a result of, e.g., capture strategy, probe type and library characteristics. Even for an entirely selective hybridization step, this presents a potential limitation for target accessibility, coverage uniformity and the fraction of target-related data from a sequencing run (Fig. 1, see below).

Another aspect of the methods is their speed and ease of automation to streamline sequencing workflows. This becomes increasingly important for larger facilities like genome centers as the number of instruments grows. Automation can for example be promoted by technical features like a simple overall workflow that avoids extensive manipulations of capture probes or sequencing libraries, use of standard steps that can be integrated into liquid handling systems or the availability of taylored hardware for automated processing. Avoiding manual intervention should also be beneficial for reproducibility, contamination risk and cost. For improvement of process speed, hybridization times seem to be a major concern since these have by far been the most time-consuming steps in previously reported methods.

Solution phase hybridization

Beside PCR and long-range PCR, several methods have been described that make use of solution phase hybridization to target sequences. Two general strategies have been developed so far. One uses different types of circularizing probes, enzymatic manipulation and generic amplification to obtain the enriched sequences [17,21,22]. The other uses enzymatically generated, long RNA probes that are immobilized on beads after hybridization for washing and elution of the desired target fragments [18].

Circularizing probes are known for applications such as FISH or SNP genotyping and have now been developed further for genome-wide sequence capture. In one method, the basic strategy is the use of 70 bp DNA “molecular inversion probes” (MIP) bearing two terminal target recognition sequences that are connected by a common linker [17,21,22]. These are generated by flexible *in situ* DNA microarray synthesis and enzymatic processing (Fig. 2). Both recognition sequences hybridize to the target loci to form a gap of ~60–190 bp that is subsequently filled in by a DNA polymerase. The resulting nick is closed by a DNA ligase and non-circularized probes are removed by an exonuclease digest, resulting in a circular library of target loci copies. An alternative approach using so-called Selector probes [17] relies on a similar probe type of 80 bp that however uses a doublestranded 40 bp common linker to assist ligation of the actual

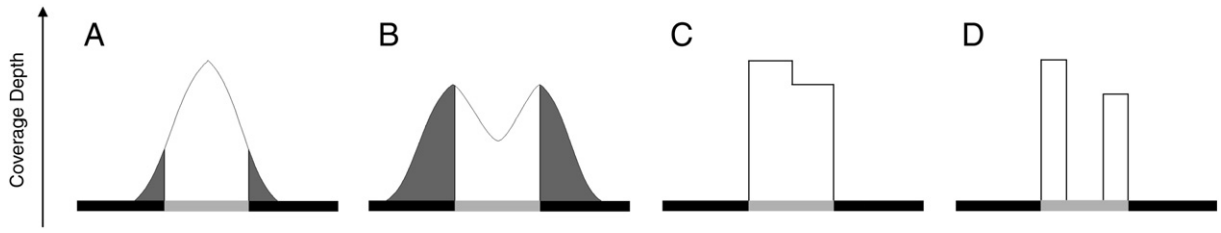

Fig. 1. Schematic view of effects of different sequence capture approaches on relative local coverage depth distributions around target regions. Genomic DNA contig is shown as bold black line with target region in light grey. Coverage depth distributions are shown as continuous black lines with white areas for on-target coverage and dark grey areas as off-target coverage. (A) Scheme of coverage depth distribution for sequencing of shotgun libraries as reported for microarray capture with probes <100 bp in length. Randomly fragmented shotgun NGS library results in binomial-like coverage distribution with maximum in the middle of the target region. A fraction of non-informative off-target reads is generated by fragments overlapping into flanking regions. Absolute proportion of non-informational reads likely depends on size of target region and fragment lengths of the NGS library. (B) Effect of sequence capture of shotgun libraries using long probes (170 bp) in combination with short end sequencing on local coverage depth distribution [18] (see also Fig. 3). Stringent hybridization selects for fragments that contain a substantial proportion of capture probe sequence. This leads to overrepresentation of fragments for which both ends map near or outside the target region boundaries. Fragments generating end sequencing reads near the middle of the target are underrepresented due to low overlap with capture probes during hybridization. This leads to a dip in the middle of the target region and diminishes the fraction of on-target reads. (C and D) Schematic views of coverage depth distributions as reported for sequence capture using molecular inversion probes (MIP, see also Fig. 2). Both ends of target fragments are fixed, resulting in a non-shotgun library. Direct sequencing leads to an even coverage of redundant reads with identical starting points for small regions where read lengths span the whole target (C). In cases where read lengths are not sufficient to span the target region, middle part is not covered (D). Varying coverage depths for both ends of the regions result from the specific setup of paired end adaptor introduction and sequencing.

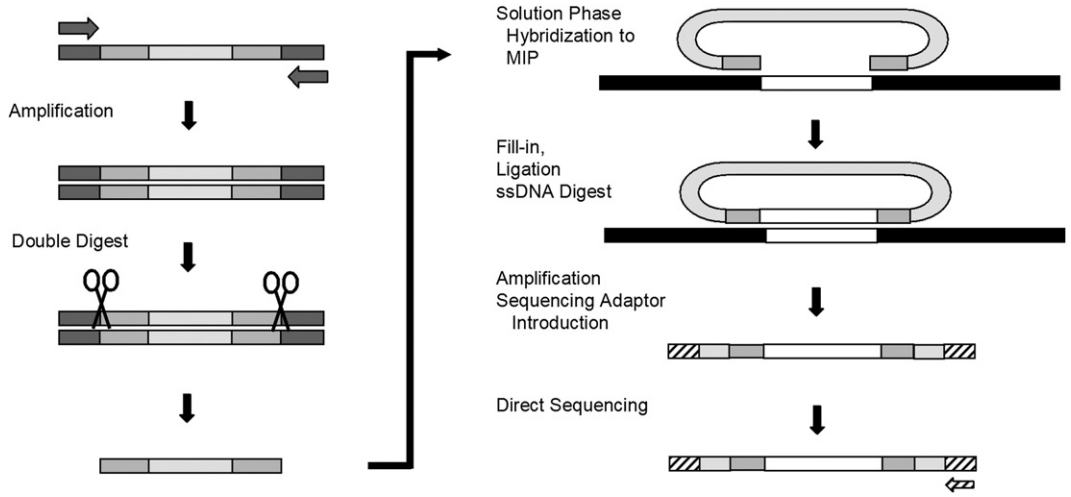

Fig. 2. Solution phase sequence capture using molecular inversion probes (MIP). Top left: A library of 100mer DNA oligonucleotides generated by microarray synthesis is designed that contains a common internal linker sequence of 30 bp (light grey), two target-specific binding regions of 20 bp each (grey) and two primer binding sequences of 15 bp each (dark grey). The library is amplified by PCR and double digested with nicking restriction endonucleases resulting in a library of 70mer ssDNA capture probes. Probes are hybridized to genomic DNA (black) and targeted regions of 60–191 bp (white) are copied in a gap-filling reaction using DNA polymerase. The resulting nick is closed by DNA ligation and non-circularized probes are selectively digested with an exonuclease. Resulting library of circular probes containing target sequence copies can be processed in two ways. For preparation of a shotgun library (not shown), a common linker-mediated PCR is performed, products are concatenated, resheared and adaptors for sequencing are attached. Alternatively, adaptors for sequencing can be attached during common linker-mediated PCR (see figure), resulting in a target library with non-random, fixed starting points for sequencing.

target DNA fragment, again resulting in a circular library. In both methods, libraries are subsequently processed and amplified to allow for introduction of adaptors for sequencing.

MIP sequence capture has been demonstrated for Illumina sequencing in two studies and with high multiplexing, targeting 13,000 or 55,000 exon targets of 100–191 bp. After a first study exhibiting poor consensus coverage and uniformity, optimization of several process steps resulted in substantial improvements. 91–98% of targets were detectably captured with no apparent allelic bias. Coverage uniformity also improved, with up to ~58% of targets being in a 10-fold range of coverage depth which is however still inferior to reported microarray capture uniformities.

Selector probe capture has so far been demonstrated in only one study with comparably low multiplexing grade. Ten genes covering 177 exons were captured and sequenced by Roche/454 sequencing. 93% of target sequence was covered by at least one read, however, this outcome required experimental testing of probes to reduce target dropout.

The overall workflows of both Selector and MIP-based methods generally have several unique characteristics with implications for capture performance. For example, there are two ways of generating the final, adaptor-ligated sequencing library after capture. In case of MIP, an elegant way has been demonstrated by primer-directed introduction of adaptors during PCR using the circular library (Fig. 2). However, this does not result in a shotgun library, i.e., start- and endpoints of the targets are fixed. Since all NGS technologies work by end-sequencing starting from NGS adaptors that are attached to the library fragments, the startpoints of all sequencing reads of a locus are identical. The size of target loci therefore becomes critical, and in the many cases where locus sizes exceed the achievable read lengths, only terminal regions are accessible for sequencing for a given probe (Fig. 1) [22].

This is a general difference between PCR- and circular probe-based approaches and methods relying on selective hybridization and recovery of a shotgun DNA library that has random start points for sequencing and thus allows for a comprehensive and more balanced coverage. The drawback can be circumvented by generating a shotgun library after capture, however, for the price of further extensive processing such as amplification, concatenation, reshearing and gel-based size selection. Since only traditional handling steps are used with MIP and Selector probes, automation would in principle be a strength of these approaches, however, the latter two steps represent a severe bottleneck for an automated workflow.

A second difference of MIP capture and PCR to solely hybridization-based methods is the fact that recognition sequences of the probes rather than the corresponding bound target sequences are part of the generated NGS reads. Consequently, to analyze the respective regions of the sample, recognition regions have to be designed adjacent to the actual targeted regions with avoiding that either one of the binding sequences targets repetitive sequence to ensure high specificity [21,22]. This is a limitation for random access to desired targets and complicates probe design in the many cases where flanking regions are low complexity sequence. Finally, all reported methods using PCR or circularizing probes introduce long noninformative sequence stretches from common linkers and/or target recognition sequences into the NGS reads. This limits data output and complicates bioinformatic analysis. This is especially true when shotgun libraries are used, since common PCR linkers contaminate reads, which lowers the mappable data amount [22].

Solution phase enrichment for Illumina NGS technology with long, biotinylated RNA probes has recently been reported [18]. Similar to MIP-based capture, initial DNA probes were synthesized with a flexible microarray platform. Each probe was 200 bp in length with a 170-bp core of target specific sequence. Probes are amplified by PCR and *in vitro* transcribed with integrated biotin-labeling (Fig. 3). Hybridization to an Illumina NGS shotgun library is then performed in solution and mixture is bound to streptavidin beads, washed to remove undesired library fragments and eluted. The method has been applied for enrichment of both an exon set and contiguous regions 0.22–2.5 Mb in size in combination with Illumina sequencing. Enrichment factors of several hundred-fold could be deduced from these studies with good coverage depths at least for the contiguous regions. Analysis of SNP calling accuracy did not hint at introduction of allelic bias.

However, a unique characteristic of the method, the very long capture probes, resulted in a drawback. Because the probe length exceeds the average length of human exons (120 bp), the maximal achievable fraction of target reads obtained from an NGS run is

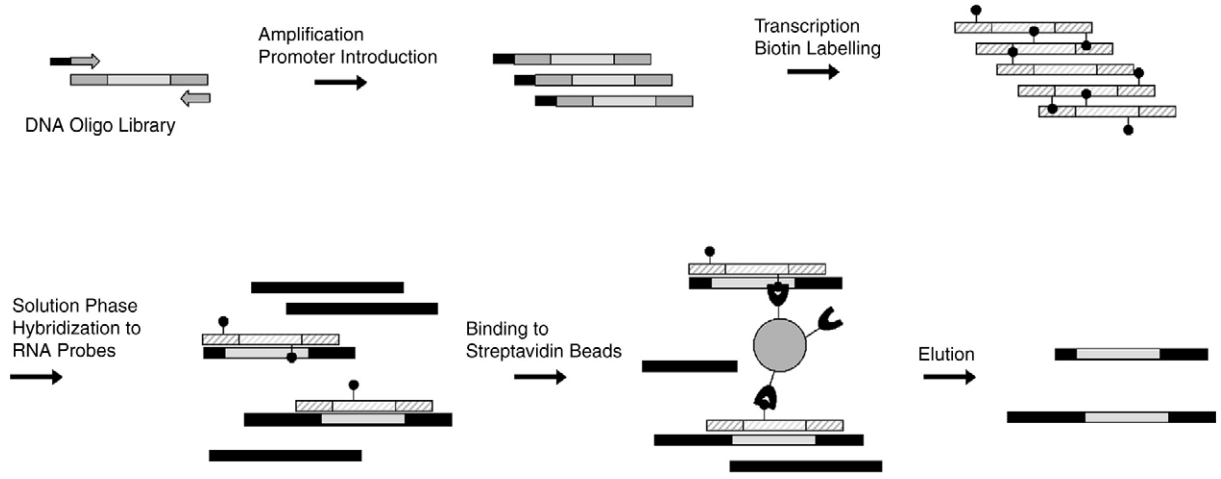

Fig. 3. Solution phase sequence capture using long RNA probes. A library of 200mer DNA oligonucleotides generated by microarray synthesis is designed that contains a target-specific core region of 170 bp (light grey) and two primer binding sequences of 15 bp each (grey). The library is amplified by PCR with introduction of a T7 promoter. *In vitro* transcription in presence of biotin-UTP results in a randomly labeled ssRNA capture probe library of 170 bp. Probes and target genomic DNA are hybridized and mixture is incubated with streptavidin beads to capture probe-target hybrids. Beads are washed and target fragments are eluted for sequencing.

intrinsically limited: Since library fragments preferentially hybridize with a maximal part of the probe sequence, a large fraction of fragment termini overlaps into flanking regions. This lead to a bimodal shape of coverage distribution and a high fraction (>50%) of off-target reads which limits overall sequencing yield for desired targets [18] (Fig. 1). Longer read lengths or construction of shotgun libraries after capture diminish this effect, however, for the price of a more complicated workflow and/or higher cost.

The approach has the advantage that all steps after array synthesis are standard procedures that can be automated on liquid handling systems for high sample throughputs. However, the generation of RNA probes from microarray oligonucleotides involving multistep enzymatic processing including PCR and *in vitro* transcription complicates the overall process in comparison to direct array-based enrichment. This in turn complicates automation and might introduce bias and errors into the probe library with potential consequences for capture specificity. Finally, hybridization times of several days had to be employed in the procedure. Given the generally favorable kinetics of solution phase hybridization, this is surprisingly long, even compared to solid phase enrichment.

Solid phase hybridization

Microarray hybridization was the first published approach for genome-wide sequence enrichment in general and has been the only platform for solid-phase enrichment so far [14-16,19,20]. However, DNA microarrays have also been used in most solution-phase enrichment approaches as source for probe precursors that are processed before the capture step. Hence, direct use of arrays for sequence enrichment results in short and very simple workflows compared to solution phase capturing (Fig. 4). Genomic DNA or a sequencing library are hybridized, the array is washed and captured DNA fragments are eluted for sequencing, optionally after library construction when genomic DNA is used.

Only two formats of arrays have been used for targeted NGS to date, both allowing for *in situ* synthesis of capture probes and thus providing high flexibility in terms of targeted sequences. This has been proven by several studies targeting both contiguous as well as exonic sequence sets. A key difference of microarray-based capture (and solution phase capture using RNA probes) compared to MIP or PCR-based approaches is that hybridized library molecules are not

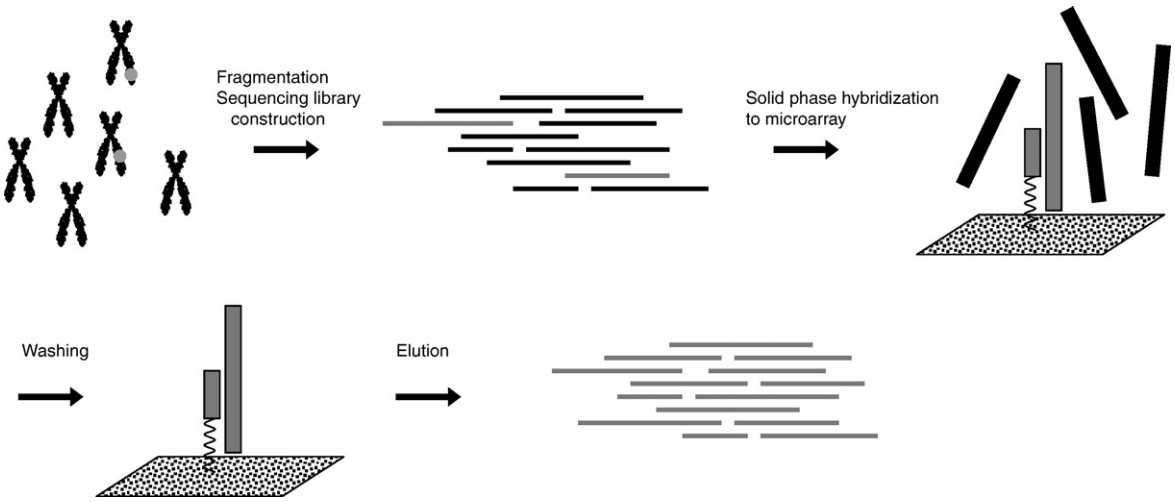

Fig. 4. Solid phase sequence capture using DNA microarrays. A shotgun library for next-generation sequencing (or fragmented genomic DNA) is hybridized to a microarray containing capture probes for targeted regions. The microarray is washed to remove unwanted sequences and residual library fragments are eluted for sequencing either directly or after construction of a shotgun sequencing library.

processed, i.e., copied or attached to target recognition sequences of capture probes (see above). This facilitates probe design since capture probes can be designed for any region within the actual target sequence.

One microarray format, a traditional glass slide array, has been used for enrichment in combination with the Roche/454 and the Illumina platforms (Fig. 4) [14,19]. DNA probes with variable lengths >60 bp were employed in a tiling array design to capture various targets up to several megabases in size. In combination with Illumina NGS, a set of all human coding exons and their adjacent splice sites representing ~1% of the genome were captured [19]. An average enrichment factor of 237-fold had been achieved, with 25% of all targeted bases covered. A study combining the same enrichment approach with Roche/454 sequencing [14] reported improved performance. Exonic or contiguous regions of 200 kb up to several Mb were captured with an average enrichment factor of 432-fold for the exonic design, a percentage of covered target bases was not reported. Analysis of reference single nucleotide polymorphism (SNP) positions did not reveal introduction of bias, which indicates applicability to resequencing studies. However, in contrast to the workflow with Illumina NGS, the employed method required introduction of artificial linkers for adaptor PCR after enrichment that were incorporated into generated sequencing reads. Additionally, the workflow had included a very long hybridization step and required multiple manual and time-consuming operations and thus, compared to solution phase methods had a relatively low potential for facile automation.

The second reported format employed compartmentalized microfluidic biochips containing eight individual array channels that allow for scalability of target sizes and/or sample numbers [15,16]. The microfluidic workflow thereby enabled the use of processing hardware featuring an almost fully automated workflow with low requirement for manual intervention. Combined with a hybridization time of only 16 h, this resulted in very short overall process times. The method used 50mer DNA probes in tiling arrays to capture several cancer-related genes for Illumina NGS in one study. Additionally, 1000 regions of 500 bp containing reference SNP positions were enriched on part of a biochip. Enrichment factors of >1000-fold have been reported with 97% or more of target region being covered. Local coverage distribution (see also Fig. 1) was reported to have a similar coverage shape as observed for enrichment using traditional microarrays with a fraction of 81% of local coverage from on-target regions. An analysis of SNP calling accuracy and allelic balance of capture indicated good performance [16]. However, similar to all other enrichment approaches, uniformity and completeness of coverage needs further optimization to allow for reliable nucleotide calling over the entire target regions.

Conclusion and outlook

Several different strategies of genomic-scale sequence enrichment have been reported with all of them having the potential to enable efficient targeted sequencing in different types of applications.

However, some current deficits are common to all methods and remain to be resolved. Variations of capture uniformity of different regions have been reported for all methods that result in dropout of difficult regions. As a consequence, none of the methods so far provided sufficient uniformity or efficiency of capture to allow for full coverage of the targeted regions at a depth allowing for reliable nucleotide calling. This seems to be a main challenge of sequence enrichment and it is not clear, if optimization of the most obvious parameters like probe design algorithms or hybridization conditions will lead to sufficient improvements. Some of the methods have already been evaluated for detection of SNPs and exhibited good performance. However, it will be interesting to also evaluate their individual applicabilities for the potentially more challenging analyses of insertions and deletions of different sizes as well as rearrangements or copy number variations. Improvements in NGS workflows might also help to overcome challenges of sequence enrichment. Longer sequencing read lengths will likely improve uniformity of coverage and could help to better cover difficult regions. As NGS technologies become more and more mature, the development of multiplexing strategies and increased sample throughput will require massively parallel enrichment of samples on a routine basis. This demands simple workflows and a maximum of automation resulting in highly standardized processes. A further requirement for standardization could arise from future incorporation of sequence enrichment technologies into diagnostic procedures.

Taken together, genomic-scale sequence enrichment is still in its infancy and a quickly developing, wide open field. It will be interesting to see, which capture concepts will prevail for which specific applications and what new and unexpected demands emerge for sequence enrichment as NGS technologies and related applications develop further. We believe that the combination of enrichment methods and high-throughput sequencing and their further improvements will transform genomics studies to the point of the creation of completely new research areas.

References

[1] D.R. Bentley, Whole-genome re-sequencing, Curr. Opin. Genet Dev. 16 (2006) 545–552.

[2] T.D. Harris, P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J.W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S.R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss, Z. Xie, Single-molecule DNA sequencing of a viral genome, Science 320 (2008) 106–109.

[3] D.S. Johnson, A. Mortazavi, R.M. Myers, B. Wold, Genome-wide mapping of in vivo protein-DNA interactions, Science 316 (2007) 1497–1502.

[4] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376–380.

[5] J. Shendure, R.D. Mitra, C. Varma, G.M. Church, Advanced sequencing technologies: methods and goals, Nat. Rev. Genet. 5 (2004) 335–344.

[6] J. Shendure, G.J. Porreca, N.B. Reppas, X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, G.M. Church, Accurate multiplex polony sequencing of an evolved bacterial genome, Science 309 (2005) 1728–1732.

[7] K. Garber, Fixing the front end, Nat. Biotechnol. 26 (2008) 1101–1104.

[8] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, J.M. Boutell, J. Bryant, R.J. Carter, R. Keira Cheetham, A.J. Cox, D.J. Ellis, M.R. Flatbush, N.A. Gormley, S.J. Humphray, L.J. Irving, M.S. Karbelashvili, S.M. Kirk, H. Li, X. Liu, K.S. Maisinger, L.J. Murray, B. Obradovic, T. Ost, M.L. Parkinson, M.R. Pratt, I.M. Rasolonjatovo, M.T. Reed, R. Rigatti, C. Rodighiero, M.T. Ross, A. Sabot, S.V. Sankar, A. Scally, G.P. Schroth, M.E. Smith, V.P. Smith, A. Spiridou, P.E. Torrance, S.S. Tzonev, E.H. Vermaas, K. Walter, X. Wu, L. Zhang, M.D. Alam, C. Anastasi, I.C. Aniebo, D.M. Bailey, I.R. Bancarz, S. Banerjee, S.G. Barbour, P.A. Baybayan, V.A. Benoit, K.F. Benson, C. Bevis, P.J. Black, A. Boodhun, J.S. Brennan, J.A. Bridgham, R.C. Brown, A.A. Brown, D.H. Buermann, A.A. Bundu, J.C. Burrows, N.P. Carter, N. Castillo, E.C.M. Chiara, S. Chang, R. Neil Cooley, N.R. Crake, O.O. Dada, K.D. Diakoumakos, B. Dominguez-Fernandez, D.J. Earnshaw, U.C. Egbujor, D.W. Elmore, S.S. Etchin, M.R. Ewan, M. Fedurco, L.J. Fraser, K.V. Fuentes Fajardo, W. Scott Furey, D. George, K.J. Gietzen, C.P. Goddard, G.S. Golda, P.A. Granieri, D.E. Green, D.L. Gustafson, N.F. Hansen, K. Harnish, C.D. Haudenschild, N.I. Heyer, M.M. Hims, J.T. Ho, A.M. Horgan, et al., Accurate whole human genome sequencing using reversible terminator chemistry, Nature 456 (2008) 53–59.

[9] J. Wang, W. Wang, R. Li, Y. Li, G. Tian, L. Goodman, W. Fan, J. Zhang, J. Li, J. Zhang, Y. Guo, B. Feng, H. Li, Y. Lu, X. Fang, H. Liang, Z. Du, D. Li, Y. Zhao, Y. Hu, Z. Yang, H. Zheng, I. Hellmann, M. Inouye, J. Pool, X. Yi, J. Zhao, J. Duan, Y. Zhou, J. Qin, L. Ma, G. Li, Z. Yang, G. Zhang, B. Yang, C. Yu, F. Liang, W. Li, S. Li, D. Li, P. Ni, J. Ruan, Q. Li, H. Zhu, D. Liu, Z. Lu, N. Li, G. Guo, J. Zhang, J. Ye, L. Fang, Q. Hao, Q. Chen, Y. Liang, Y. Su, A. San, C. Ping, S. Yang, F. Chen, L. Li, K. Zhou, H. Zheng, Y. Ren, L. Yang, Y. Gao, G. Yang, Z. Li, X. Feng, K. Kristiansen, G.K. Wong, R. Nielsen, R. Durbin, L. Bolund, X. Zhang, S. Li, H. Yang, J. Wang, The diploid genome sequence of an Asian individual, Nature 456 (2008) 60–65.

[10] H. Wolinsky, The thousand-dollar genome. Genetic brinkmanship or personalized medicine, EMBO Rep. 8 (2007) 900–903.

[11] M. Meyer, U. Stenzel, S. Myles, K. Prufer, M. Hofreiter, Targeted high-throughput sequencing of tagged nucleic acid samples, Nucleic Acids Res. 35 (2007) e97.

[12] Y. Erlich, K. Chang, A. Gordon, R. Ronen, O. Navon, M. Rooks, G.J. Hannon, DNA Sudoku–harnessing high-throughput sequencing for multiplexed specimen analysis, Genome Res. (2009).
[13] A. Kahvejian, J. Quackenbush, J.F. Thompson, What would you do if you could sequence everything? Nat. Biotechnol. 26 (2008) 1125–1133.
[14] T.J. Albert, M.N. Molla, D.M. Muzny, L. Nazareth, D. Wheeler, X. Song, T.A. Richmond, C.M. Middle, M.J. Rodesch, C.J. Packard, G.M. Weinstock, R.A. Gibbs, Direct selection of human genomic loci by microarray hybridization, Nat. Methods 4 (2007) 903–905.
[15] S. Bau, N. Schracke, M. Kranzle, H. Wu, P.F. Stahler, J.D. Hoheisel, M. Beier, D. Summerer, Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays, Anal. Bioanal. Chem. 393 (2009) 171–175.
[16] D. Summerer, H. Wu, B. Haase, Y. Cheng, N. Schracke, C.F. Staehler, M.S. Chee, P.F. Stahler, M. Beier, Microarray-based Multicycle-Enrichment of Genomic Subsets for Targeted Next-Generation-Sequencing, Genome Res. 19 (9) (2009) 1616–1621.
[17] F. Dahl, J. Stenberg, S. Fredriksson, K. Welch, M. Zhang, M. Nilsson, D. Bicknell, W.F. Bodmer, R.W. Davis, H. Ji, Multigene amplification and massively parallel sequencing for cancer mutation discovery, Proc. Natl. Acad Sci. U. S. A. 104 (2007) 9387–9392.
[18] A. Gnirke, A. Melnikov, J. Maguire, P. Rogov, E.M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D.B. Jaffe, E.S. Lander, C. Nusbaum, Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing, Nat. Biotechnol. 27 (2009) 182–189.
[19] E. Hodges, Z. Xuan, V. Balija, M. Kramer, M.N. Molla, S.W. Smith, C.M. Middle, M.J. Rodesch, T.J. Albert, G.J. Hannon, W.R. McCombie, Genome-wide in situ exon capture for selective resequencing, Nat. Genet 39 (2007) 1522–1527.
[20] D.T. Okou, K.M. Steinberg, C. Middle, D.J. Cutler, T.J. Albert, M.E. Zwick, Microarray-based genomic selection for high-throughput resequencing, Nat. Methods 4 (2007) 907–909.
[21] G.J. Porreca, K. Zhang, J.B. Li, B. Xie, D. Austin, S.L. Vassallo, E.M. LeProust, B.J. Peck, C.J. Emig, F. Dahl, Y. Gao, G.M. Church, J. Shendure, Multiplex amplification of large sets of human exons, Nat. Methods 4 (2007) 931–936.
[22] E.H. Turner, C. Lee, S.B. Ng, D.A. Nickerson, J. Shendure, Massively parallel exon capture and library-free resequencing across 16 genomes, Nat. Methods 6 (2009) 315–316.